DOCUMENT RESUME

ED 445 148 UD 033 745

AUTHOR Plank, Stephen B.; Young, Estelle
TITLE Lessons for Scaling Up: Evaluations of the Talent Development Middle School's Student Team Literature Program.
INSTITUTION Center for Research on the Education of Students Placed At Risk, Baltimore, MD.
SPONS AGENCY Office of Educational Research and Improvement (ED), Washington, DC.; Carnegie Corp. of New York, NY.
REPORT NO CRESPAR-R-46
PUB DATE 2000-06-00
NOTE 35p.
CONTRACT R-117-D40005
AVAILABLE FROM For full text: http://www.csos.jhu.edu.
PUB TYPE Reports - Research (143)
EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Educational Change; Educational Policy; *Middle School Students; Middle Schools; Models; Peer Teaching; Program Evaluation; *Program Implementation; Reading Comprehension; *Talent; *Urban Schools
IDENTIFIERS *Philadelphia School District PA

ABSTRACT

Comprehensive school reform efforts are an increasingly visible part of the educational landscape. Policymakers, educators, and researchers are eager to assess the effectiveness of these models, especially regarding their utility in the most troubled settings. This paper reports results from one such reform, the Talent Development Middle School (TDMS). Focusing on reading comprehension, the report analyzes data from two Philadelphia, Pennsylvania, middle schools that have been implementing TDMS and from two comparison schools. Hierarchical linear models suggest that TDMS has had overall positive effects on achievement. In general, otherwise similar students in Talent Development schools outperform comparison students, controlling for prior achievement. However, this overall TDMS effect is neither mediated nor accompanied by a positive effect of one recommended component of the TDMS approach -- peer-assisted learning. In addition, differential effects of Talent Development by grade were found. In the two Talent Development schools, which serve grades 5 through 8, some troubling trends have been found in sixth grade results. In light of this finding, implications for model development and future research are discussed. An appendix describes the TDMS model's core curriculum in detail. (Contains 7 tables, 2 figures, and 23 references.) (Author/SLD)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 445 148

CRESPAR

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

LESSONS FOR SCALING UP

Evaluations of the Talent Development Middle School's Student Team Literature Program

Stephen B. Plank
Estelle Young

Report No. 46 / June 2000

BEST COPY AVAILABLE

FUNDED BY
OFFICE OF EDUCATIONAL RESEARCH AND IMPROVEMENT
U.S. DEPARTMENT OF EDUCATION

JOHNS HOPKINS UNIVERSITY & HOWARD UNIVERSITY

CENTER FOR RESEARCH ON THE EDUCATION OF STUDENTS PLACED AT RISK

UD033745

LESSONS FOR SCALING UP
Evaluations of the Talent Development Middle School's Student Team Literature Program

Stephen B. Plank and Estelle Young

Johns Hopkins University

Report No. 46

June 2000

Published by the Center for Research on the Education of Students Placed At Risk (CRESPAR), supported as a national research and development center by the Office of Educational Research and Improvement (OERI), U.S. Department of Education (R-117-D40005). Additional funding for this work was received from the Carnegie Corporation of New York. The opinions expressed in this article do not necessarily reflect the positions or policies of OERI or the Carnegie Corporation, and no official endorsement should be inferred. An on-line version of this report is available at our web site: www.csos.jhu.edu.

CRESPAR TECHNICAL REPORTS

1996

1. *The Talent Development High School: Essential Components*—V. LaPoint, W. Jordan, J.M. McPartland, D.P. Towns
2. *The Talent Development High School: Early Evidence of Impact on School Climate, Attendance, and Student Promotion*—J.M. McPartland, N. Legters, W. Jordan, E.L. McDill
3. *The Talent Development Middle School: Essential Components*—S. Madhere, D.J. Mac Iver
4. *The Talent Development Middle School: Creating a Motivational Climate Conducive to Talent Development in Middle Schools: Implementation and Effects of Student Team Reading*—D.J. Mac Iver, S.B. Plank
5. *Patterns of Urban Student Mobility and Local School Reform: Technical Report*—D. Kerbow
6. *Scaling Up: Lessons Learned in the Dissemination of Success for All*—R.E. Slavin, N.A. Madden
7. *School-Family-Community Partnerships and the Academic Achievement of African American, Urban Adolescents*—M.G. Sanders
8. *Asian American Students At Risk: A Literature Review*—S.-F. Siu

1997

9. *Reducing Talent Loss: The Impact of Information, Guidance, and Actions on Postsecondary Enrollment*—S.B. Plank, W.J. Jordan
10. *Effects of Bilingual Cooperative Integrated Reading and Composition on Students Transitioning from Spanish to English Reading*—M. Calderón, R. Hertz-Lazarowitz, G. Ivory, R.E. Slavin
11. *Effective Programs for Latino Students in Elementary and Middle Schools*—O.S. Fashola, R.E. Slavin, M. Calderón, R. Duran
12. *Detracking in a Racially Mixed Urban High School*—R. Cooper
13. *Building Effective School-Family-Community Partnerships in a Large Urban School District*—M.G. Sanders
14. *Volunteer Tutoring Programs: A Review of Research on Achievement Outcomes*—B.A. Wasik
15. *Working Together to Become Proficient Readers: Early Impact of the Talent Development Middle School's Student Team Literature Program*—D.J. Mac Iver, S.B. Plank, R. Balfanz
16. *Success for All: Exploring the Technical, Normative, Political, and Socio-Cultural Dimensions of Scaling Up*—R. Cooper, R.E. Slavin, N.A. Madden
17. *MathWings: Early Indicators of Effectiveness*—N.A. Madden, R.E. Slavin, K. Simons
18. *Parental Involvement in Students' Education During Middle School and High School*—S. Catsambis, J.E. Garland

1998

19. *Success for All/Éxito Para Todos: Effects on the Reading Achievement of Students Acquiring English*—R.E. Slavin, N.A. Madden
20. *Implementing a Highly Specified Curricular, Instructional, and Organizational School Design in a High-Poverty, Urban Elementary School: Three Year Results*—B. McHugh, S. Stringfield
21. *The Talent Development Middle School: An Elective Replacement Approach to Providing Extra Help in Math—The CATAMA Program (Computer and Team-Assisted Mathematics Acceleration)*—D.J. Mac Iver, R. Balfanz, S.B. Plank
22. *School-Family-Community Partnerships in Middle And High Schools: From Theory to Practice*—M.G. Sanders, J.L. Epstein
23. *Sources of Talent Loss Among High-Achieving Poor Students*—W.J. Jordan, S.B. Plank
24. *Review of Extended-Day and After-School Programs and Their Effectiveness*—O.S. Fashola
25. *Teachers' Appraisals of Talent Development Middle School Training, Materials, and Student Progress*—E. Useem

1998 (cont'd)

26. *Exploring the Dynamics of Resilience in an Elementary School*—S.M. Nettles, E.P. Robinson
27. *Expanding Knowledge of Parental Involvement in Secondary Education: Effects on High School Academic Success*—S. Catsambis
28. *Socio-Cultural and Within-School Factors That Effect the Quality of Implementation of School-Wide Programs*—R. Cooper

1999

29. *How Students Invest Their Time Out of School: Effects on School Engagement, Perceptions of Life Chances, and Achievement*—W.J. Jordan, S.M. Nettles
30. *Disseminating Success For All: Lessons for Policy and Practice*—R.E. Slavin, N.A. Madden
31. *Small Learning Communities Meet School-to-Work: Whole-School Restructuring for Urban Comprehensive High Schools*—N.E. Legters
32. *Family Partnerships with High Schools: The Parents' Perspective*—M.G. Sanders, J.L. Epstein, L. Connors-Tadros
33. *Grade Retention: Prevalence, Timing, and Effects*—N. Karweit
34. *Preparing Educators for School-Family-Community Partnerships: Results of a National Survey of Colleges and Universities*—J.L. Epstein, M.G. Sanders, L.A. Clark
35. *How Schools Choose Externally Developed Reform Designs*—A. Datnow
36. *Roots & Wings: Effects of Whole-School Reform on Student Achievement*—R.E. Slavin, N.A. Madden
37. *Teacher Collaboration in a Restructuring Urban High School*—N.E. Legters
38. *The Child First Authority After-School Program: A Descriptive Evaluation*—O.S. Fashola
39. *MathWings: Effects on Student Mathematics Performance*—N.A. Madden, R.E. Slavin, K. Simons
40. *Core Knowledge Curriculum: Three-Year Analysis of Implementation and Effects in Five Schools*—B. McHugh, S. Stringfield
41. *Success for All/Roots & Wings: Summary of Research on Achievement Outcomes*—R.E. Slavin, N.A. Madden

2000

42. *The Role of Cultural Factors in School Relevant Cognitive Functioning: Synthesis of Findings on Cultural Contexts, Cultural Orientations, and Individual Differences*—A.W. Boykin, C.T. Bailey
43. *The Role of Cultural Factors in School Relevant Cognitive Functioning: Description of Home Environmental Factors, Cultural Orientations, and Learning Preferences*—A.W. Boykin, C.T. Bailey
44. *Classroom Cultural Ecology: The Dynamics of Classroom Life in Schools Serving Low-Income African American Children*—C.M. Ellison, A.W. Boykin, D.P. Towns, A. Stokes
45. *An "Inside" Look at Success for All: A Qualitative Study of Implementation and Teaching and Learning*—A. Datnow, M. Castellano
46. *Lessons for Scaling Up: Evaluations of the Talent Development Middle School's Student Team Literature Program*—S.B. Plank, E. Young
47. *A Two-Way Bilingual Program: Promise, Practice, and Precautions*—M. Calderón, Argelia Carreón (in press)
48. *Four Models of School Improvement: Successes and Challenges in Reforming Low-Performing, High-Poverty Title I Schools*—G.D. Borman, L. Rachuba, A. Datnow, M. Alberg, M. Mac Iver, S. Stringfield, S. Ross (in press)
49. *National Evaluation of Core Knowledge Sequence Implementation: Final Report*—S. Stringfield, A. Datnow, G. Borman, L. Rachuba (in press)
50. *Core Knowledge Curriculum: Five-Year Analysis of Implementation and Effects in Five Maryland Schools*—M.A. Mac Iver, S. Stringfield, B. McHugh (in press)

BEST COPY AVAILABLE

The Center

Every child has the capacity to succeed in school and in life. Yet far too many children, especially those from poor and minority families, are placed at risk by school practices that are based on a sorting paradigm in which some students receive high-expectations instruction while the rest are relegated to lower quality education and lower quality futures. The sorting perspective must be replaced by a "talent development" model that asserts that all children are capable of succeeding in a rich and demanding curriculum with appropriate assistance and support.

The mission of the Center for Research on the Education of Students Placed At Risk (CRESPAR) is to conduct the research, development, evaluation, and dissemination needed to transform schooling for students placed at risk. The work of the Center is guided by three central themes — ensuring the success of all students at key development points, building on students' personal and cultural assets, and scaling up effective programs — and conducted through research and development programs in the areas of early and elementary studies; middle and high school studies; school, family, and community partnerships; and systemic supports for school reform, as well as a program of institutional activities.

CRESPAR is organized as a partnership of Johns Hopkins University and Howard University, and supported by the National Institute on the Education of At-Risk Students (At-Risk Institute), one of five institutes created by the Educational Research, Development, Dissemination and Improvement Act of 1994 and located within the Office of Educational Research and Improvement (OERI) at the U.S. Department of Education. The At-Risk Institute supports a range of research and development activities designed to improve the education of students at risk of educational failure because of limited English proficiency, poverty, race, geographic location, or economic disadvantage.

Abstract

Comprehensive school reform efforts are an increasingly visible part of the educational landscape. Policymakers, educators, and researchers are eager to assess the effectiveness of these models, especially regarding their utility in the most troubled settings. We report results for one such reform, the Talent Development Middle School (TDMS). Focusing on reading comprehension, we analyze data from two Philadelphia middle schools which have been implementing TDMS, and two comparison schools. Hierarchical linear models suggest that TDMS has had overall positive effects on achievement. In general, otherwise similar students in Talent Development schools outperform comparison students, controlling for prior achievement. However, this overall TDMS effect is neither mediated nor accompanied by a positive effect of one recommended component of the TDMS approach — peer-assisted learning. Also, we find differential effects of Talent Development by grade which demand further attention. Specifically, in these two Talent Development schools which serve fifth through eighth grades, we find some troubling trends in the sixth grade results. In light of our findings, implications for model development and future research are discussed.

Acknowledgments

The authors thank Douglas Mac Iver, Robert Balfanz, Leslie Jones, and Alta Shaw for their substantive comments on this report. We also appreciate greatly the time and careful attention given by Barbara Colton and Marc Cutright in reading and editing this report.

Introduction

How great is the promise of comprehensive, whole-school reform models? Can whole-school improvement strategies developed by university-based researchers or other "outsiders" be introduced to schools in ways that lead to meaningful adoption by teachers and gains in student achievement? If successes are realized, can they be replicated across schools and over time?

Whole-school reform models are an increasingly visible part of the American educational landscape. Educators, policymakers, and the research community are eager to assess the effectiveness of these models. Inquiry is especially focused upon whether this kind of reform can work in schools challenged with large class sizes, high student mobility, and concentrated poverty (Knapp, 1995). Careful evaluations and reporting by both model developers and third-party analysts are needed.

We are part of a team of researchers and educators at Johns Hopkins University that has been developing and studying one of these school improvement models, the Talent Development Middle School (TDMS). The TDMS pilot sites are several high-poverty, urban schools. In this report, we attempt to evaluate the effects of the TDMS model and, more specifically, the Student Team Literature approach to teaching English and language arts.

The TDMS model is intended to establish the curriculum, instruction, school organization, and professional development needed in order for all students in a middle school to learn challenging academic materials and prepare for successful futures. The eight key components of the model are: (1) a demanding core curriculum, (2) detracking of instruction, (3) facilitated, standards-based instructional programs, (4) extra help, enrichment, and recognition, (5) communal organization, (6) school-family-community partnerships, (7) cultural relevance, and (8) career and educational exploration (Mac Iver, Balfanz, & Plank, 1998; Mac Iver et al., in press; Mac Iver & Plank, 1997).

As we are using the phrase "comprehensive, whole-school reform model," we mean a set of practices for curriculum, instruction, organization, and professional development that is intended to affect virtually all subject areas, classrooms, and students in a school in pursuit of more equitable and effective opportunities to learn, and improved student performance. Currently, numerous comprehensive reform efforts are being attempted and evaluated (American Institutes for Research, 1999; Bodilly, 1998; Hatch, 1998; Stringfield, Ross, & Smith, 1996). These various efforts differ somewhat in the extent to which they attempt, or are able, to manipulate various school practices and policies. In TDMS efforts to date, our partnerships with schools and districts have allowed us to make recommendations and offer support in the areas of curricular materials, instructional practices, professional development,

and some aspects of school organization (e.g., encouraging small learning communities, looping, and extra-help opportunities in core subjects). But while we have made recommendations in each of these areas, we have not ultimately had the power to mandate or enforce implementation. Furthermore, we have not been able to alter factors such as the length of the school day or year, the hiring or firing of teachers, or — for the most part — class size.

The first school to adopt the TDMS model was Central East Middle School in Philadelphia, which began implementation in 1995. By Spring 1999, four more Philadelphia public schools had adopted the model. For research purposes, these Philadelphia pilot sites have been paired with demographically matched comparison schools in the same school district. Additionally, a national field test is now being initiated in a new set of schools in Philadelphia, Detroit, and Memphis.

Student Team Literature (STLit), the cornerstone of the TDMS approach to teaching English and language arts, was among the first parts of the model to be developed and implemented. STLit includes (1) curricular materials, (2) recommended instructional practices, peer assistance processes and assessments, and (3) professional development, mentoring, and advising to support these curricular and instructional reforms (Mac Iver & Plank, 1996, 1997; Mac Iver, Plank, & Balfanz, 1997). The curricular materials are centered on grade-appropriate, award-winning novels. The novels are supplemented with detailed teacher guides and student materials that facilitate teacher-led instruction, peer-assisted learning, and regular assessments that directly align with classroom activities. STLit is designed to be particularly effective in promoting higher-order thinking and reading comprehension skills. It is complemented by Talent Development Writing and Listening Comprehension activities.[1]

Central to the instructional practices and professional development of STLit are cooperative learning and peer-assisted activities. A body of prior research and educational theory suggests that peer-assisted learning can have positive effects on student motivation, effort, peer support for achievement, and achievement (Cohen, 1994; Cohen & Lotan, 1997; Fuchs et al., 1997; King, 1994; Mevarech & Kramarski, 1997; Slavin, 1995; Webb & Farivar, 1994). Peer-assisted learning can also allow for more frequent feedback to students on their academic progress. In light of this prior research and theory, STLit, and specifically peer-assisted learning opportunities, can be expected to increase student achievement in reading comprehension and other areas of language arts achievement.

It is important to note that peer-assisted learning is just one component of cooperative learning, as the latter is usually defined. Cooperative learning refers to a variety of teaching methods in which students work in small groups to help each other master academic subjects and skills. Effective cooperative learning methods incorporate group goals and individual

accountability (Cohen, 1994; Slavin, 1995). In contrast to traditional pedagogy in which only the teacher supervises and instructs, cooperative learning involves a shifting of authority from the teacher to the students, who become largely responsible for their teammates' effort and understanding during team study sessions. The methods are responsive to the developmental needs of middle school students, as they allow for peer interaction and offer opportunities for self-direction and autonomy. Further, cooperative learning is designed to shift the focus of the inevitable peer pressure and interpersonal comparisons of adolescence in positive directions which support effort and academic achievement.

As researchers have begun to examine the conditions under which cooperative learning works best, it has been shown that the effectiveness of cooperative learning is related to the time and effort dedicated to preparing students for the cooperative experience and for their roles as peer tutors to the other members of their teams. Although middle school students usually respond favorably to opportunities to interact with peers during learning activities, they often lack the social skills and strategies for resolving conflicts, staying on task, and involving all teammates (Williams, Harris, & Hayakawa, 1995). To realize the greatest possible benefits of cooperative learning, it is important to help students develop skills in basic communication, conflict resolution, peer tutoring, and task completion (Fuchs et al., 1994; Meloth & Deering, 1994; Webb & Farivar, 1994).

In this report we measure and estimate the effects of peer-assisted learning, not the broader construct of cooperative learning. As such, our results should not be taken as a test of the effectiveness of cooperative learning. Rather, our analyses are an attempt to investigate further some effects of STLit and the TDMS model, holistically, that we detected in previous research (Mac Iver, Plank, & Balfanz, 1997). Additionally, that previous research revealed important patterns related to peer-assisted learning.

Specifically, our prior research on the effects of STLit in the first year of implementation (1995-96) at Central East Middle School (CEMS) indicated that (1) peer assistance was occurring more frequently at CEMS than at its matched comparison school; (2) CEMS students' reading comprehension levels were significantly higher than those of comparison students, controlling for prior achievement and grade level; and (3) the effects of peer assistance on achievement differed by school (Mac Iver, Plank, & Balfanz, 1997). Regarding the differential effects by school, we found some suggestion that peer assistance may have had a positive effect at CEMS but a negative effect at the comparison school.[2] We interpreted this difference as indicating a qualitative difference in the ways students and teachers interacted during small group learning times. We noted that CEMS staff had much exposure to professional development and teaching materials through their participation in the Talent Development program.

Table 1
School Characteristics: Central East and Cooke and Comparison, 1996-1997

	Central East	Comparison	Cooke	Comparison
Student Characteristics				
Total Number of Students	1040	1005	1118	860
Race/Ethnicity				
Black	25.7%	19.9%	72.3%	72.0%
Hispanic	45.6%	60.9%	9.4%	0.9%
White	14.1%	18.5%	1.3%	9.4%
Other	14.6%	0.7%	17.0%	17.7%
Percent Low Income	85.9%	95.7%	86.8%	90.6%
Special Education	11.5%	16.8%	10.1%	9.3%
Limited English Proficient	7.8%	8.1%	9.6%	4.1%
Staff Characteristics				
Total Staff	88	98	101	85
Administrative	10.2%	7.1%	7.9%	7.1%
Teaching	65.9%	70.4%	70.3%	65.9%
Student/Teacher Ratio	18.9	15.2	16.2	15.9
Percent Female	80.7%	69.4%	69.6%	71.8%
Race/Ethnicity				
Black	38.4%	34.0%	52.5%	44.7%
Hispanic	5.8%	6.2%	2.0%	0.0%
White	54.7%	59.8%	41.6%	55.3%
Other	1.2%	0.0%	4.0%	0.0%

Note: Administrative staff are defined as the principal, assistant principal, and clerical staff. Teaching staff are defined as classroom teachers, special education teachers, and teacher's aides. For the purposes of calculating the student/teacher ratio, teacher's aides were excluded. Early childhood staff were also excluded from the staff count. From the School District of Philadelphia, http://www2.phila.k12.pa.us.

We now have available two more years of achievement and survey data from CEMS and its comparison site. We also have first-year data from the second school to adopt the Talent Development Middle School model. This school, Cooke Middle School in Philadelphia, began implementation of the model in Fall 1997. For Cooke and a matched comparison site, we have achievement data from October 1997, and April 1998, as well as survey data describing students' experiences during the 1997-98 school year. With these new data, we have the opportunity to compare and contrast findings from these additional years and schools with the results for CEMS in 1995-96.

All four schools — CEMS, Cooke, and their comparison sites — are non-selective, public schools serving the fifth, sixth, seventh, and eighth grades. The comparison sites were selected by the research office of the School District of Philadelphia, based on similarities to CEMS and Cooke in terms of school size, grade span, prior achievement, student demographics, and teacher characteristics. Table 1 shows characteristics of each school's students and staff during the 1996-97 school year. Each school is quite diverse racially and ethnically. The socioeconomic levels of the schools' constituent families are quite low, with at least 85 percent being categorized as low income. Table 1 is intended to show that the Talent Development and comparison schools are well-matched. Later tables present comparisons of the schools' achievement levels. These levels, too, are matched closely enough to allow for a quasi-experimental research design in which prior achievement is controlled in models predicting end-of-year achievement.

Research Questions

The main research questions addressed in this report are:

1. Overall, is there evidence that Talent Development schools facilitate significantly greater growth in reading comprehension than do comparison schools?
2. If an effect is detected, is this effect partially explained by students' exposure to peer-assisted learning activities?
3. Is a general pattern of results consistently found across grades, across successive years of implementation, and across schools?

Data

The data for this study come from a larger multi-disciplinary evaluation of the Talent Development Middle School program. The evaluation draws on standardized test scores, student surveys of classroom experience, ethnographic observation, focus groups of teachers, and interviews with students and teachers. For this report, we rely on two of these sources, standardized test scores and student surveys. Both are administered annually to all students in all grades at Talent Development and comparison schools. Prior to the implementation of Talent Development, standardized tests were administered to generate baseline achievement data. For Central East, we are not able to include fifth graders in the analyses, as we do not have test data from the previous spring.

Measure of Achievement: Reading Comprehension. We use standardized test scores from the Stanford 9 to capture achievement because of their tested reliability and the availability of national norms that permit comparisons. Reading comprehension was operationalized as a student's reading comprehension scale score from the Stanford 9's multiple choice battery. For CEMS and its comparison site, we have test score data from April of 1996, 1997, and 1998. For Cooke and its comparison site, we have test data from October 1997 and April 1998. Test score results come in various metrics such as raw scores and normal curve equivalents. The advantage of the scale score metric is that the calibration is done at the interval level and is constant across time and grade. Thus the distance between scale score units can be considered to be equivalent and enables absolute comparisons across grade and year. This is not true of the other metrics.

Tables 2, 3, and 4 contain descriptive statistics of test scores for CEMS and its comparison site for a sample of students with test scores in years 1996-97 and 1997-98 and Cooke and its comparison site in Fall 1997 and Spring 1998. The corresponding grade-specific normal curve equivalent and grade equivalent is reported alongside the scale scores to enable comparison relative to national norms for grade-level performance (NCE=50). These tables are discussed later in the report.

Measure of Conformity to the TDMS Model: Peer Assistance. While the Talent Development model is a comprehensive set of organizational, professional development, curricular, and pedagogical reforms, our statistical models in this report include only one specific dimension of this reform — peer assistance. Peer assistance is meant to capture part of the larger pedagogical reform of cooperative learning, in which mixed-ability pairs or groups of students work together in a structured setting. Peer assistance items were included in a questionnaire, administered to all students, that asked them to reflect on their experiences in English/language arts class during that school year.

Specifically, peer assistance is measured by a four-item scale measuring the frequency of peer assistance and discussion in English/language arts class. By computing the mean student response to these items within each classroom, we were able to create a scale that reliably measured the frequency with which peer assistance activities had occurred during the school year. These items, following a prompt that asked how often various things had happened in language arts class, were: (1) students took turns with partners asking questions and answering the questions the partners asked; (2) students discussed a novel with a partner; (3) students worked in teams to master the vocabulary used in a novel; and (4) students explained answers to their teammates and checked to make sure that all their teammates understood the material.[3] These surveys were administered at CEMS and its comparison site in February or March of 1997 and 1998. They were administered at Cooke and its comparison site in May 1998.

Table 2

Test Score Descriptives: Cooke and Comparison, Fall 1997-Spring 1998

	Cooke		Comparison	
SAT-9 Reading Comprehension	Fall 1997	Spring 1998	Fall 1997	Spring 1998
Grade 5				
Scale Scores	581.44	626.59	597.98	620.58
Grade Equivalent	3.08	4.69	3.55	4.36
Normal Curve Equiv	19.04	37.61	26.30	34.82
	(*n*=112)		(*n*=185)	
Grade 6				
Scale Scores	606.63	624.95	620.36	634.82
Grade Equivalent	3.77	4.61	4.42	5.18
Normal Curve Equiv	24.61	31.48	31.16	36.85
	(*n*=183)		(*n*=174)	
Grade 7				
Scale Scores	620.43	646.22	630.26	648.88
Grade Equivalent	4.53	5.95	4.90	6.05
Normal Curve Equiv	24.85	35.21	29.56	36.51
	(*n*=195)		(*n*=157)	
Grade 8				
Scale Scores	647.08	669.94	651.02	666.61
Grade Equivalent	5.98	7.63	6.17	7.30
Normal Curve Equiv	32.68	40.84	34.25	39.13
	(*n*=167)		(*n*=135)	
% of classrooms scoring at least 2 grades below grade level	86%	48%	63%	41%
% of students scoring at least 2 grades below grade level	70%	56%	58%	53%
% of classrooms scoring at most 1 grade below grade level	7%	7%	4%	0%
% of students scoring at most 1 grade below grade level	19%	30%	24%	31%

Table 3

Test Score Descriptives: Central East and Comparison, Spring 1997-Spring 1998

	Central East		Comparison	
SAT-9 Reading Comprehension	Spring 1997	Spring 1998	Spring 1997	Spring 1998
Grade 6				
Scale Scores	634.59	637.09	614.14	623.83
Grade Equivalent	5.07	5.33	4.05	4.56
Normal Curve Equiv	40.98	37.89	30.42	31.22
	(*n*=141)		(*n*=77)	
Grade 7				
Scale Scores	623.64	666.28	623.76	641.74
Grade Equivalent	4.64	7.39	4.54	5.57
Normal Curve Equiv	30.02	45.45	29.76	32.83
	(*n*=171)		(*n*=144)	
Grade 8				
Scale Scores	654.48	670.25	646.04	664.38
Grade Equivalent	6.56	7.63	5.76	7.21
Normal Curve Equiv	39.58	40.93	34.94	37.99
	(*n*=174)		(*n*=188)	
% of classrooms scoring at least 2 grades below grade level	24%	14%	60%	64%
% of students scoring at least 2 grades below grade level	52%	45%	66%	58%
% of classrooms scoring at most 1 grade below grade level	43%	43%	16%	28%
% of students scoring at most 1 grade below grade level	36%	42%	24%	26%

Table 4

Test Score Descriptives: Central East and Comparison, Spring 1996-Spring 1997

	Central East		Comparison	
SAT-9 Reading Comprehension	Spring 1996	Spring 1997	Spring 1996	Spring 1997
Grade 6				
Scale Scores	622.03	631.72	612.79	622.83
Grade Equivalent	4.46	4.95	3.96	4.42
Normal Curve Equiv	36.02	34.39	31.12	29.34
	(*n*=106)		(*n*=42)	
Grade 7				
Scale Scores	637.81	658.68	624.43	645.63
Grade Equivalent	5.38	6.79	4.49	5.71
Normal Curve Equiv	38.72	41.63	31.82	34.70
	(*n*=128)		(*n*=128)	
Grade 8				
Scale Scores	660.49	675.01	643.51	660.68
Grade Equivalent	6.99	8.12	5.63	6.80
Normal Curve Equiv	42.54	43.46	33.74	36.16
	(*n*=148)		(*n*=145)	
% of classrooms scoring at least 2 grades below grade level	16%	16%	73%	77%
% of students scoring at least 2 grades below grade level	50%	46%	70%	62%
% of classrooms scoring at most 1 grade below grade level	37%	37%	12%	23%
% of students scoring at most 1 grade below grade level	38%	40%	19%	28%

Methods

Our main analyses consist of two-level hierarchical linear models in which students (level-1) are nested within language arts classrooms (level-2). For each pair of schools and each testing time point, we begin by estimating a one-way ANOVA model to establish a baseline estimate of the proportion of total variance residing within and between classrooms. From that baseline model, we build a set of nested models.[4] The final models for CEMS and Comparison A in 1997, CEMS and Comparison A in 1998, and Cooke and Comparison B in 1998 differ somewhat from one another, according to which blocks of variables and error components improved the empirical fit of each model. An example of the mathematical notation for the most complex of our final models — that for Cooke/Comparison B in 1998 — is as follows:

$$Y_{ij} = \beta_{0j} + \beta_{1j} (\text{Prior Achievement}_{ij}) + r_{ij}$$

$$\beta_{0j} = \gamma_{00} + \gamma_{01} (\text{Cooke}_j) + \gamma_{02} (6^{\text{th}}\ \text{grade}_j) + \gamma_{03} (7^{\text{th}}\ \text{grade}_j) + \gamma_{04} (8^{\text{th}}\ \text{grade}_j) + \gamma_{05} (\text{Peer Assistance}_j) + \gamma_{06} (\text{Cooke}_j * 6^{\text{th}}\ \text{grade}_j) + \gamma_{07} (\text{Cooke}_j * 7^{\text{th}}\ \text{grade}_j) + \gamma_{08} (\text{Cooke}_j * 8^{\text{th}}\ \text{grade}_j) + \gamma_{09} (\text{Cooke}_j * \text{Peer Assistance}_j) + u_{0j}$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11} (\text{Cooke}_j) + \gamma_{12} (6^{\text{th}}\ \text{grade}_j) + \gamma_{13} (7^{\text{th}}\ \text{grade}_j) + \gamma_{14} (8^{\text{th}}\ \text{grade}_j) + \gamma_{15} (\text{Cooke}_j * 6^{\text{th}}\ \text{grade}_j) + \gamma_{16} (\text{Cooke}_j * 7^{\text{th}}\ \text{grade}_j) + \gamma_{16} (\text{Cooke}_j * 8^{\text{th}}\ \text{grade}_j) + u_{1j},$$

where prior achievement is grand mean-centered, all other predictors are uncentered, and the usual assumptions are made about error terms (Bryk & Raudenbush, 1992).

We employ hierarchical linear models (HLM) to investigate our research questions because HLM explicitly models the nested nature of students clustered together within classrooms. In the technical sense, this allows one to cast aside the dubious assumption that the achievement of a student is independent from other students in his or her classroom. HLM permits an explicit test to see if there are *classroom-level differences* in achievement. In the equations above, each language arts classroom is denoted by the subscript j; the model indicates that each classroom is associated with its own slope and intercept. Significant variation among intercept coefficients indicates classroom-level differences in the predicted level of achievement. Variation among slope coefficients reflects the extent to which classroom context affects the rate of change in estimated achievement. These classroom-specific parameters in HLM are subsequently modeled in separate regressions, called level-2 regressions, using variables at the classroom level of analysis (such as mean classroom peer assistance) to estimate the classroom-level slope and intercept. In the present analyses, a level-2 dummy variable for school and a level-2 measure of peer assistance are used to evaluate the effects of TDMS, globally, and a particular mode of instruction, specifically.

Results

The findings are presented to answer the research questions posed in the beginning of this report. To simplify the discussion, we use a single model, that of Cooke and its comparison site in 1997-98, to answer the first two questions in detail. We will draw on the other two years of data, as well as previously published results, to compare and contrast findings across years, grades, and schools.

As a backdrop to the ensuing discussion, we first refer to Table 2 with test scores for 1997-98 for Cooke and its comparison site. Scores for 1996-97 and 1997-98 for CEMS and its comparison site are included in Tables 3 and 4 as a reference. The top portion of Table 2 shows the average reading comprehension test scores by grade and school in Fall 1997 and Spring 1998 for those students with test scores in both semesters. Three metrics are displayed — scale scores, grade equivalents, and normal curve equivalents. The fall scores are considered baseline scores as the first full year of implementation of Student Team Literature at Cooke occurred during the 1997-98 school year. Looking first at the fall scores, Cooke students perform significantly *below* the comparison site students. This is true across all grades, but the gap is greatest in the fifth and sixth grades. By the spring, however, Cooke has closed the gap or even passed comparison site students in all grades, although least so in the sixth grade. Despite these gains, however, both schools still lag significantly behind national norms in Spring 1998.

The bottom portion of the table focuses on another aspect of performance by highlighting students and classrooms performing above or below a particular standard. In schools with the demographic composition of Cooke and its comparison site, students are often two years behind grade level by the time they reach middle school. For this reason, we used a benchmark of two or more years below grade level as a standard to mark the proportion of students and classrooms considered underperforming. We defined students and classrooms as performing up to standard if their test scores were within a year of grade level. By looking at students *and* classrooms, one can get a sense of the uniformity of the distribution of low- and high-performing students across classrooms. For example, the proportion of high-performing students at Cooke rose from 19 percent to 30 percent between Fall 1997 and Spring 1998. This improvement is an overall school improvement, and not relegated to a particular set of classrooms within Cooke, since the proportion of *classrooms* whose average student scored within this high-performing range did not change between the fall and spring. Instead, the improvement in student performance meant a sharp drop in the proportion of underperforming classrooms, from 86 percent to 48 percent, indicating that enough students in those classrooms improved their scores sufficiently to make the overall class average improve.

Table 5 shows the results for the models of 1998 achievement at Cooke and its comparison site. The table contains the parameter estimates for the level-2 regressions for the slope and intercept models as well as the variance components. In the body of this report, we will discuss in detail the full model for Cooke and discuss the models for Central East in less detail. Because the final models differ, we also include a reduced form model, with only prior achievement, school, and grade as explanatory variables to enable comparison across schools and time-points.

We begin the discussion of Cooke using the reduced model. While not statistically significant at the .05 level ($p<.07$), Cooke students were estimated to achieve approximately 5 scale score points above comparison site students, controlling for prior achievement and grade. Turning to the more complex model, which includes peer assistance and the interactions of school with grade and peer assistance in the model of the intercept, the overall school effect is grade-dependent. This is seen by examining the interaction terms between school and grade, as well as the main effect of school. The biggest school advantage for the Talent Development school occurs in the fifth grade. Cooke fifth graders (with prior achievement at the grand mean for both schools) were estimated to outperform comparison fifth graders by 16 scale score points (15.97), controlling for peer assistance. This advantage is almost as great in the eighth grade (15.97 - 3.90 = 12.07), attenuated significantly in the seventh grade (15.97 - 9.96 = 6.01), and reversed in the sixth grade (15.97 - 18.61 = -2.64).[5] Note that the model facilitates comparisons of *estimated achievement* between Talent Development and comparison students with *the same prior achievement score*, and thus differs from the picture in the descriptive table. The descriptive table shows that Cooke students, in practice, start with lower prior achievement scores than the comparison site.

The model for Cooke also includes estimates for the prior achievement slope and reveals a negative but insignificant difference by school for fifth graders, where prior achievement has a smaller effect at Cooke than at the comparison site [(.50 - .04 = .46) and (.50), respectively]. In the sixth grade, the relationship between prior achievement and end-of-year achievement at Cooke is approximately one-third lower than that observed at the comparison school [(.50 + .16 - .04 - .20 = .42) and (.50 +.16 = .66), respectively]. This finding of a flatter slope at Cooke in some grades can be interpreted as implying a relative disadvantage for high achievers at Cooke compared to the matched site. It adds legitimacy to a concern that high-achieving students' learning might be slowed under reforms such as Talent Development where students are grouped in heterogeneous classrooms.[6] However, since Cooke had relatively few students achieving above the two schools' grand mean in Fall 1997 (72%, 73%, 75%, and 61% of fifth, sixth, seventh, and eighth graders, respectively, scored at least two grades below grade level), the majority of students at Cooke benefitted from the combination of the higher intercept but flatter slope at Cooke.

Table 5

Peer Assistance, School and Grade Level Effects on Student Reading Comprehension Achievement: Cooke and Comparison, 1998 (HLM Estimates)

	School Model: Intercept as Outcome (School and Grade Only)			Full Model: Intercept and Slope as Outcomes (School, Grade and Peer Assistance)		
Fixed Effects	Coefficient	*se*	*p*	Coefficient	*se*	*p*
Model for Level-1 intercept, β_{0j}						
Intercept, γ_{00}	636.83	3.02	0.00	630.16	3.90	0.00
Cooke, γ_{01}	5.03	2.71	0.07	15.97	6.84	0.02
Sixth Grade, γ_{02}	-7.15	3.77	0.06	1.92	5.25	0.72
Seventh Grade, γ_{03}	4.23	3.84	0.28	9.94	5.36	0.07
Eighth Grade, γ_{04}	11.92	4.17	0.01	13.19	6.17	0.04
Freq. of Peer Assistance, γ_{05}				-8.33	7.45	0.27
Cooke * Sixth Grade, γ_{06}				-18.61	8.44	0.03
Cooke * Seventh Grade, γ_{07}				-9.96	8.28	0.24
Cooke * Eighth Grade, γ_{08}				-3.90	8.80	0.66
Cooke * Freq. of Peer Assist., γ_{09}				3.84	9.28	0.68
Model for prior achievement slope, β_{1j}						
Intercept, γ_{10}	0.54	0.02	0.00	0.50	0.05	0.00
Cooke, γ_{11}				-0.04	0.07	0.55
Sixth, γ_{12}				0.16	0.07	0.03
Seventh, γ_{13}				0.05	0.07	0.51
Eighth, γ_{14}				0.12	0.08	0.17
Cooke * Sixth, γ_{15}				-0.20	0.11	0.07
Cooke * Seventh, γ_{16}				0.11	0.11	0.32
Cooke * Eighth, γ_{17}				0.00	0.12	0.97

Random Effects	Variance Component	*df*	χ^2	*p*	Variance Component	*df*	χ^2	*p*
Classroom mean, u_{0j}	75.06	50	154.50	0.00	68.70	46	168.71	0.00
Prior achievement slope, u_{1j}	.0026	54	72.44	0.05				
Level-1 effect, r_{ij}	577.54				576.40			

Note: Proportion of variance in adjusted classroom mean reading comprehension achievement explained by level-2 variables: 45%=(124.13-68.70)/124.13. Proportion of within classroom variance in reading achievement explained by prior achievement: 36%=(906.22-577.68)/906.22.

Table 6

Peer Assistance, School and Grade Level Effects on Student Reading Comprehension Achievement: Central East Middle School and Comparison, 1998 (HLM Estimates)

	School Model: Intercept as Outcome (School and Grade Only)			Full Model: Intercept as Outcome (School, Grade and Peer Assistance)		
Fixed Effects	Coefficient	*se*	*p*	Coefficient	*se*	*p*
Model for Level-1 intercept, β_{0j}						
Intercept, γ_{00}	633.23	3.85	0.00	631.09	4.91	0.00
CEMS, γ_{01}	7.88	3.64	0.04	6.26	6.36	0.33
Seventh Grade, γ_{02}	20.35	4.56	0.00	13.78	6.08	0.03
Eighth Grade, γ_{03}	20.65	4.48	0.00	24.26	5.95	0.00
Freq. of Peer Assistance, γ_{04}				-16.47	6.89	0.02
CEMS* Seventh Grade, γ_{05}				17.93	8.42	0.04
CEMS * Eighth Grade, γ_{06}				-5.23	8.51	0.54
CEMS * Freq. of Peer Assist., γ_{07}				12.90	10.65	0.23
Model for prior achievement slope, β_{1j}						
Intercept, γ_{10}	0.62	0.04	0.00	0.62	0.04	0.00

Random Effects	Variance Component	*df*	χ^2	*p*	Variance Component	*df*	χ^2	*p*
Classroom mean, u_{0j}	118.09	41	169.95	0.00	84.03	37	119.00	.00
Prior achievement slope, u_{1j}	0.0324	44	96.76	0.00	.0412	44	97.60	.00
Level-1 effect, r_{ij}	588.19				583.25			

Note: Proportion of variance in adjusted classroom mean reading comprehension achievement explained by level-2 variables: 60%=(212.15-84.03)/212.15. Proportion of within classroom variance in reading achievement explained by prior achievement: 42%=(1013.03-586.56)/1013.03.

Table 7

Peer Assistance, School and Grade Level Effects on Student Reading Comprehension Achievement: Central East Middle School and Comparison, 1997 (HLM Estimates)

	School Model: Intercept as Outcome (School and Grade Only)			Full Model: Intercept and Slope as Outcomes (School, Grade and Peer Assistance)		
Fixed Effects	Coefficient	*se*	*p*	Coefficient	*se*	*p*
Model for Level-1 intercept, β_{0j}						
Intercept, γ_{00}	638.15	4.30	0.00	627.38	5.10	0.00
CEMS, γ_{01}	-0.20	4.01	0.96	7.98	4.98	0.12
Seventh Grade, γ_{02}	13.00	4.74	0.01	20.07	5.85	0.00
Eighth Grade, γ_{03}	17.33	5.03	0.00	23.19	6.06	0.00
Freq. of Peer Assistance, γ_{04}				-18.83	9.30	0.05
CEMS * Freq. of Peer Assist., γ_{05}				14.67	12.69	0.26
Model for prior achievement slope, β_{1j}						
Intercept, γ_{10}	0.58	0.04	0.00	0.34	0.08	0.00
CEMS, γ_{11}				0.14	0.07	0.05
Seventh Grade, γ_{12}				0.20	0.09	0.03
Eighth Grade, γ_{13}				0.21	0.09	0.02

Random Effects	Variance Component	*df*	χ^2	*p*	Variance Component	*df*	χ^2	*p*
Classroom mean, u_{0j}	237.55	43	333.47	0.00	202.65	41	284.50	.00
Prior achievement slope, u_{1j}	.0244	46	72.91	0.01	.0163	43	56.63	.08[a]
Level-1 effect, r_{ij}	495.82				494.92			

[a] A likelihood ratio test of the difference in deviance between a fixed and random specification indicates a significant improvement in the goodness of fit with a random effects specification: χ^2=6896-6885=11, *df*=2, $p<.01$.

Note: Proportion of variance in adjusted classroom mean reading comprehension achievement explained by level two variables: 35%=(312.64-202.65)/312.64. Proportion of variance in slope of prior achievement explained by level-2 variables: 35%=(.0252-.0163)/.0252. Proportion of within-classroom variance in reading achievement explained by prior achievement: 43%=(864.55-495.79)/864.55.

While the above discussion indicates an overall school effect and shows an advantage to students at the Talent Development school, this school difference alone does not necessarily demonstrate any positive impact of Talent Development per se, but could arise from other school-level differences having little to do with Talent Development. We can eliminate some competing reasons for the school effect, thereby narrowing the possible explanations. For example, Table 1 established that Cooke and its comparison school have very similar student demographic compositions. Also, Cooke enjoys no advantages over the comparison that we can detect in terms of average class size, general qualifications of staff, desirability of assignment to the school, or stability of leadership.

Although we cannot be certain that the overall school effect is attributable to the TDMS model, it is instructive to pursue a sort of component analysis to see whether one particular aspect of Talent Development accounts for some of the school-level difference. We perform this analysis by turning our attention to peer assistance, a construct explicitly related to Student Team Literature. As seen in Table 5, peer assistance has a negative, but insignificant, main effect on the intercept, implying that an increase in the student-reported level of peer assistance slightly (non-significantly) reduces predicted achievement scores, controlling for prior achievement, grade, and school. Considering both the main effect of peer assistance and the interaction term of peer assistance with Cooke (-8.33 and +3.84, respectively), the negative effect of peer assistance is attenuated but not reversed at Cooke. This finding runs counter to the intent of the Talent Development model which employs cooperative learning practices as a means of increasing student motivation, effort, and achievement.

Another way to describe our findings is to make predictions, based on the estimated model, for hypothetical students with conceptually relevant values of the independent variables. Figures 1 and 2 illustrate the expected outcomes for students at the Talent Development and comparison schools. Figure 1 plots estimated 1998 scale scores for students in classrooms with grand mean peer assistance across a range of prior achievement. The solid lines depict the almost universal advantage of Cooke students across all grade levels and prior test results, with the notable exception of the sixth grade where only low prior achievers are expected to perform better at Cooke than the comparison site. This disadvantage for sixth graders with high prior achievement stems from the negative interaction terms of school and sixth grade in the intercept and slope; comparison site students do best relative to Cooke students in the sixth grade if they are moderate or high prior achievers. Under no other scenario would a student at the comparison school be expected to outperform a similarly situated Cooke student. Thus the Cooke advantage is quite robust. The Cooke advantage is most pronounced in the fifth grade, followed by the eighth and seventh grades. In the seventh grade case, however, the Cooke advantage is most pronounced for high prior achievers.

Figure 1. Predicted 1998 Reading Comprehension: Cooke and Comparison School, with Grand-Mean Peer Assistance

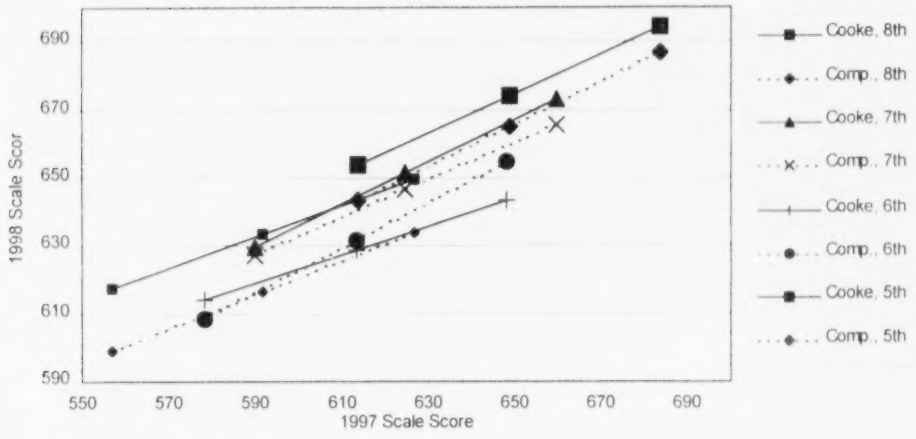
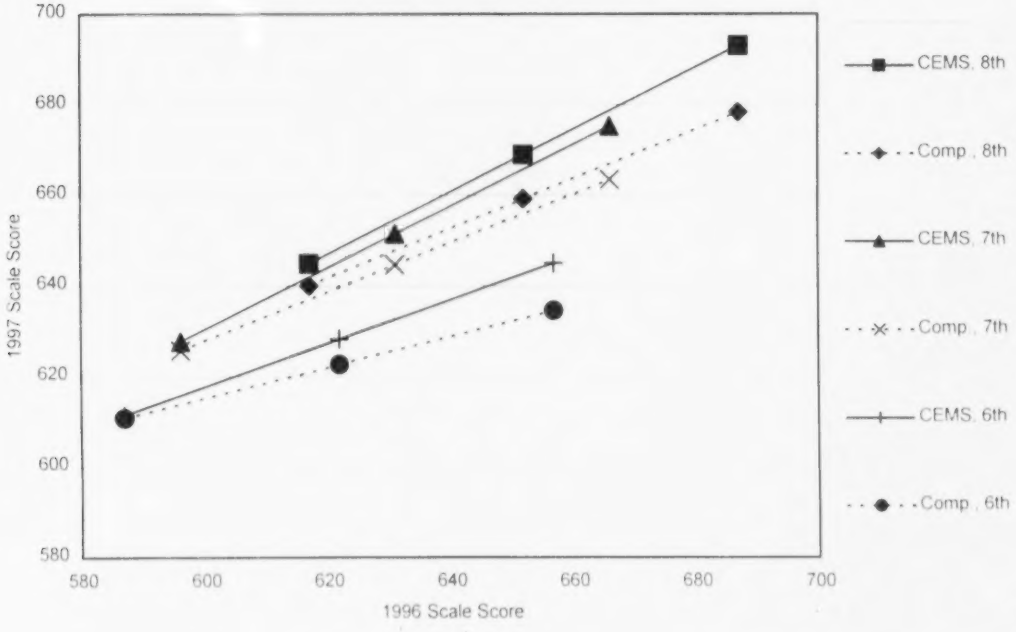


Figure 2. Predicted 1997 Reading Comprehension: CEMS and Comparison School, with Grand-Mean Peer Assistance



In order to see how generalizable these findings are, we can compare the findings from Cooke with findings from Central East, both those presented in this report and a previous one (Mac Iver, Plank, & Balfanz, 1997). Although the exact composition of the best-fitting models differ by school and time-point, several patterns emerge. Two points are somewhat troubling in their implications for the actual implementation of cooperative learning and whole-school reform. First, peer assistance has an overall negative and substantively large main effect that is attenuated, but generally not reversed, in Talent Development schools. In only one case, that of previously published findings from the first year of implementation in Central East, was the effect of peer assistance positive at the Talent Development school. Second, the significant coefficients for grade and the interactions between school and grade suggest that the TDMS model is not uniformly effective across all grades, despite its intended whole-school nature. At Cooke, sixth and seventh graders hold the smallest advantage relative to comparison students. At Central East, sixth grade is the locus of least advantage. In noting variations by grade, however, we do not want to lose sight of a larger point — there are overall advantages associated with attending a Talent Development school.[7] These advantages hold for all students at all levels of prior achievement at all time-points at both Talent Development schools except for moderate and high prior achieving sixth graders at Cooke.

The models also contain some unique features indicating processes which are not consistent across schools and/or time-points. In two of the four models (Central East in 1996 and 1997), a positive Talent Development school effect on the prior achievement slope was found, whereas a negative (though insignificant) Talent Development school effect on this slope was found for Cooke. For these two years at Central East, students with high prior achievement scores were estimated to gain the most from attending a Talent Development school.

Discussion and Conclusion

There is fairly consistent evidence that Talent Development schools significantly improve reading comprehension. This finding is repeated across schools and time points — otherwise similar students in Talent Development schools outperform comparison students. However, we have not found any evidence that the overall Talent Development effect is either mediated or accompanied by a positive influence of peer assistance, the one specific indicator of Talent Development implementation modeled in this report. It seems likely, though, that particular components of the Talent Development model, in particular the high-quality, comprehensive curricular materials and the professional development activities, are making a positive difference even if other elements of the model are unevenly or poorly implemented.

While we suspect that the curricular materials, professional development activities, and recommended instructional practices of TDMS are central to the model's positive effects on reading comprehension, there are many aspects of these for which we have not systematically collected data. Some aspects of these are difficult to measure, particularly through the student surveys that have been our main source of information about classroom activities and environments, to date. For example, while student surveys have provided an apparently valid and reliable measure of peer assistance, it is more difficult to document the degree to which a teacher proceeds through a Partner Discussion Guide in the recommended fashion.

Now, though, as our focus falls increasingly on curriculum, instruction, and professional development, future studies should investigate the breadth, scope, and quality of implementation of these components of the Talent Development model. As an initial step, we administered a survey to all 96 English/language arts teachers in the five current Talent Development schools in June 1999, asking about their experiences implementing Student Team Literature during the 1998-99 school year. The survey includes items about exposure to professional development activities such as annual training workshops in Student Team Literature, and classroom visits by instructional facilitators or curriculum coaches. The survey also includes items on the use of the Partner Discussion Guides, the comprehensive curricular material developed for each reading selection in Talent Development schools. In addition, the survey includes items on a number of other aspects of the Talent Development model. Additionally, external evaluators have begun conducting studies, including focus groups and in-depth interviews with students, teachers, and administrators (Useem, 1998; Wilson & Corbett, 1999). Together these data will provide a more comprehensive picture of the quality of implementation of Talent Development than that which can be gleaned from student surveys alone. These data will also complement the existing data from the student surveys and might help clarify and illuminate the confusing findings about peer assistance.

We offer a few other caveats and comments as we conclude this report. Although we have revealed some negative or null relationships between peer-assisted learning and achievement, we feel strongly that it would be wrong to take this as evidence against the effectiveness of cooperative learning. Our current study leads us to conclude that something is happening in the Talent Development schools to prompt greater gains in reading comprehension than is seen at demographically similar schools in the same district. High levels of peer-assisted learning do not explain this advantage. However, peer-assisted learning may be occurring in many Talent Development classrooms (and many non-Talent Development classrooms) without some other important components of cooperative learning. In particular, we suspect that the social skill development that must precede and accompany effective cooperative learning is deficient. Additionally, the absence of a positive effect for

peer-assisted learning could stem from the failure of TDMS facilitators and teachers to become proficient in assuring that time dedicated to group or partner work is on-task, productive learning time for students.

Our findings of differential effects by grade, particularly that the between-school differences are small or non-existent in the sixth grade, implies a poorly performing sixth grade relative to the other years in the Talent Development schools. This finding resonates with the painful dilemma faced by too many troubled school districts. In a middle school serving four grades, the earliest grade often gets special consideration as a school tries to nurture the youngest students making the transition from elementary school, and understandably so. And with the importance of the transition to high school, and accountability systems that many districts have in place that focus on the final year or two of middle school, the final one or two grades also get special consideration. Just as an emergency room doctor makes decisions based on triage and must leave some patients unattended, principals find that they must decide how to allocate their most talented teachers, an action that by definition leaves those classrooms not designated with high priority status with the weakest teachers. The sobering reality is that these schools do not enjoy an unlimited supply of exceptional teachers, and so must allocate those teachers according to these priorities. While we can hope that they nonetheless work to improve the skills of their weaker teachers, the pattern of our findings is consistent with the implication of this dilemma of resource allocation.

The Talent Development model shows much promise, but has yet to reach its potential until we can confidently state that all teachers have been reached and all classrooms influenced. These findings speak to the myriad challenges of implementation of whole-school reform initiatives, particularly in high-poverty areas. The goals now are to continue the capacity building within the Talent Development schools and to gain a more thorough understanding of their processes and outcomes through research.

Endnotes

1. Additional details about STLit, Talent Development Writing, and Listening Comprehension are provided in the appendix.

2. These findings are suggestive at best as the p-values associated with a main effect of peer assistance and an interaction term of peer assistance by CEMS were 0.147 and 0.093, respectively.

3. To measure how often each learning activity occurred in each classroom, we computed the mean response (in a z-score metric) of all the students in the class. First, students' responses were coded in terms of school days per month (i.e., never=0 days per month, once or twice a month=1.5 days per month, once or twice a week=6 days per month, most days=12 days per month, every day=20 days per month). Z-scores were then calculated for each of the items. Then, the classroom mean z-score for each item was computed. Finally, a scale score for each class was computed by averaging classroom mean z-scores for the four items in the scale.

4. There were two specification tests performed. For the level-2 equations, a multi-parameter test to compare the nested models with and without a block of theoretically related variables (e.g. school and grade) was performed using a likelihood ratio test of the difference in deviance between the less and more restricted models. This multi-parameter test was performed using full maximum likelihood estimation to compute the deviance (Bryk and Raudenbush, pp. 44-46). Selection between random and fixed effect specifications of the level-2 equations was made based on a likelihood ratio test of the difference in deviance. If the random effects model did not significantly improve the goodness of fit, the fixed effect model was estimated. Once a model was selected, however, the actual estimates of the coefficients were obtained using restricted maximum likelihood, as full maximum likelihood assumes equal group size for unbiased estimates of sigma-squared. Restricted maximum likelihood allows unequal group size and small sample size and still yields unbiased estimates.

5. Note that only the interaction term of school and sixth grade is significant in the model of the intercept and that it is only marginally significant in the model of the slope ($p<.07$). The likelihood ratio test of the difference in deviance indicated a significant improvement in the goodness of fit for the inclusion of the block of interaction terms in the slope but not for the intercept. However, we retained the interaction terms in both the slope and the intercept since it made little theoretic sense to include them in the slope but exclude them in the level-2 model of the intercept.

6. In our previous analyses (Mac Iver, Plank, & Balfanz, 1997), we used the logic that the combination of a higher intercept and steeper prior achievement slope at CEMS implied positive effects for all CEMS students, but especially for those with high prior achievement. Thus, we concluded that fears that high achieving students would suffer in heterogeneous classrooms should be allayed.

7. A further way to express this overall advantage is in terms of an effect size, which can be calculated by dividing the school coefficient from the reduced model (see Tables 5, 6, and 7) by the between-classroom standard deviation of the reading comprehension measure. This standard deviation is obtained from a one-way ANOVA model (not shown). For the schools analyzed in this paper, these effect sizes are 0.51 for CEMS in 1995-96, -0.01 for CEMS in 1996-97, 0.38 for CEMS in 1997-98, and 0.24 for Cooke in 1997-98. For the near-zero effect in 1996-97, we offer the caveat that a positive and significant effect emerges in the intercept and slope as outcomes model; a positive effect seems to be masked until the complexities of differing slopes by grade and school are taken into account. Thus, overall, we feel there is a positive Talent Development effect even in this year.

References

American Institutes for Research (1999). *An educator's guide to schoolwide reform.* Arlington, VA: Educational Research Service.

Bodilly, S. (1998). *Lessons from New American Schools' scale-up phase: Prospects for bringing designs to multiple schools.* Santa Monica, CA: Rand.

Bryk, A.S., & Raudenbush, S.W. (1992). *Hierarchical linear models: Applications and data analysis methods.* Newbury Park, CA: Sage Publications.

Cohen, E.G. (1994). *Designing groupwork. Strategies for the heterogeneous classroom.* New York: Teachers College Press.

Cohen, E.G., & Lotan, R.A. (Eds.) (1997). *Working for equity in heterogeneous classrooms: Sociological theory in practice.* New York: Teachers College Press.

Fuchs, D., Fuchs, L.S., Mathes, P.G., & Simmons, D.C. (1997). Peer-assisted learning strategies: Making classrooms more responsive to diversity. *American Educational Research Journal, 34*(1): 174-206.

Fuchs, L.S., Fuchs, D., Bentz, J., Phillips, N.B., & Hamlett, C.L. (1994). The nature of student interactions during peer tutoring with and without prior training and experience. *American Educational Research Journal, 31*(1): 75-103.

Hatch, T. (1998). The differences in theory that matter in the practice of school improvement. *American Educational Research Journal, 35*(1): 3-31.

King, A. (1994). Guiding knowledge construction in the classroom: Effects of teaching children how to question and how to explain. *American Educational Research Journal, 31*(2): 338-368.

Knapp, M.S. (1995). *Teaching for meaning in high-poverty classrooms.* New York: Teachers College Press.

Mac Iver, D.J., Balfanz, R., & Plank, S. (1998). An "elective replacement" approach to providing extra help in math: The Talent Development Middle Schools' Computer- and Team-Assisted Mathematics Acceleration (CATAMA) program. *Research in Middle Level Education Quarterly* 22(2): 1-23.

Mac Iver, D.J., Mac Iver, M.A., Balfanz, R., Plank, S.B., & Ruby, A. (in press). Talent Development Middle Schools: Blueprint and results for a comprehensive whole-school reform model. In M. Sanders (Ed.), *Schooling students placed at risk: Research, policy,*

and practice in the education of poor and minority adolescents. Mahwah, NJ: Lawrence Erlbaum Associates.

Mac Iver, D.J., & Plank, S.B. (1996). *Creating a motivational climate conducive to talent development in middle schools: Implementation and effects of Student Team Reading*. (Report 4). Baltimore, MD, & Washington, DC: Center for Research on the Education of Students Placed At Risk.

Mac Iver, D.J., & Plank, S.B. (1997). Improving urban schools: Developing the talents of students placed at risk. In J.L. Irvin (Ed.), *What current research says to the middle level practitioner* (pp.243-256). Columbus, OH: National Middle School Association.

Mac Iver, D.J., Plank, S.B., & Balfanz, R. (1997). *Working together to become proficient readers: Early impact of the Talent Development Middle School's Student Team Literature Program* (Report 15). Baltimore, MD, & Washington, DC: Center for Research on the Education of Students Placed At Risk.

Meloth, M.S., & Deering, P.D. (1994). Task talk and task awareness under different cooperative learning conditions. *American Educational Research Journal, 31*(1): 138-165.

Mevarech, Z.R., & Kramarski, B. (1997). IMPROVE: A multidimensional method for teaching mathematics in heterogeneous classrooms. *American Educational Research Journal, 34*(2): 365-394.

Slavin, R.E. (1995). *Cooperative learning: Theory, research, and practice*. (2nd ed.). Boston: Allyn and Bacon.

Stringfield, S., Ross, S., & Smith, L. (Eds.). (1996). *Bold plans for school restructuring: The New American Schools Designs*. Mahwah, NJ: Lawrence Erlbaum Associates.

Useem, E.L. (1998). *Teachers' appraisals of Talent Development Middle School training, materials, and student progress: Results from focus grous* (Report 20). Baltimore, MD, & Washington, DC: Center for Research on the Education of Students Placed At Risk.

Webb, N.M., & Farivar, S. (1994). Promoting helping behavior in cooperative small groups in middle school mathematics. *American Educational Research Journal, 31*(2): 369-395.

Williams, D.R., Harris, J., & Hayakawa, C. (1995). Cooperative learning and conflict resolution: Perspectives of urban middle school adolescents. *Research in Middle Level Education, 18*(2): 23-47.

Wilson, B.L. & Corbett, H.D. (1999). *"No excuses:" The eighth grade year in six Philadelphia Middle Schools*. Philadelphia: Philadelphia Education Fund.

APPENDIX

Further Description of STLit, Talent Development Writing, and Listening Comprehension

The TDMS model's core English/language arts (ELA) curriculum includes Student Team Literature, Talent Development Writing, and Listening Comprehension. These three components represent a research- and standards-based approach to teaching ELA that is designed to build both basic language skills and higher-order thinking skills and to extend reading comprehension skills. These components are integrated in the ELA curriculum to work together, rather than in isolation, to teach reading, writing, and comprehension skills. For example, during Listening Comprehension, a teacher reads aloud a passage from the novel students are currently reading. Literature Related Writing assignments, in turn, are based on the current reading selection. The teacher explicitly teaches skills through modeling, making analogies, and guided practice, and these skills are reinforced and refined throughout the year.

Student Team Literature aims to enable students to apply effective reading strategies and operations while engaged in the reading process; strengthen cognitive elaboration; extend comprehension skills; enable students to acquire and benefit from knowledge of the author's craft; help students to read, interpret, and interact with a variety of literary works to explore the human experience and develop critical appreciation; develop fluency in reading and writing; increase students' vocabularies; provide frequent opportunities to extend speaking and listening skills; and encourage cooperative work. The instructional practices and curricular material in Student Team Literature allow teachers to guide students in learning and applying skills in meaningful tasks, engaging in self-evaluation against explicitly defined expectations, and mastering higher-order thinking such as appreciating multiple forms and voices.

The principal vehicle to organize the process of meeting these goals is the Partner Discussion Guide. A Partner Discussion Guide accompanies each book, short story, non-fiction selection, or poem. The purpose of the Partner Discussion Guide is to organize the students' activities about a reading selection to improve their comprehension and give them knowledge that will help them in future reading activities of the same kind. Student Team Literature provides both teacher and student editions of the guides.

The teacher's edition of a Partner Discussion Guide contains all the necessary content for a lesson. These include reading exercises to engage students in and prepare students for the day's reading and develop and practice strategic reading skills; teaching activities to model the day's lesson and teach specific reading, writing, and critical thinking skills; cooperative learning activities including partner reading and discussion to provide the opportunity for

guided practice of skills, interpretation, and understanding of the reading selection; and prompts for reflection and review activities. Partner Discussion Guides also provide sufficient reflection, synthesis, and review opportunities so that all students have the opportunity to study and master the material during informal pre-assessment activities prior to formal graded tests. The guides contain a variety of written, spoken, and listening activities designed to be mutually reinforcing to support even a struggling reader's progress through an entire piece of literature. The teachers' Partner Discussion Guide also contains sufficient background material about the author, the book, the literary devices it contains, and other relevant information so that even a teacher previously unfamiliar with a particular piece of literature or one teaching out of his or her content area can obtain the expertise necessary to guide students through the book.

Guiding begins before the selection is read. The teacher engages students in pre-reading activities in order to activate their prior knowledge, spark their interest in the upcoming reading, and increase their confidence in their ability to understand the selection. Pre-reading activities include previewing the book, based on its title, the cover picture, the book flaps, back cover, and selected vocabulary words from the book. The teacher uses these parts of the book to help students make predictions. These activities prepare students for the ideas in a particular selection, introduce important but unfamiliar vocabulary, and enable students to better understand the selection by making connections with what they already know. Pre-reading activities are essential to learning to read strategically.

After beginning the reading selection, the teacher shares with the class *meaningful sentences* using vocabulary words from the day's reading. The teacher models the writing of meaningful sentences by writing sentences of his or her own in which a word is used in such a way that to substitute a different word would render the sentence nonsensical. This way, students have an opportunity to develop vocabulary in context. With both direct instruction in which the teacher explicitly models the composition of meaningful sentences and daily guided practice, students learn to think critically about vocabulary as well as integrate their thoughts about what is being read with those of the author.

Most of the selection is read individually, as Student Team Literature places an emphasis on silent reading. The other elements of Student Team Literature also support successful comprehension, retention, and self-monitoring during silent reading. One option, recommended for poor readers, those whose primary language is not English, resource poor classrooms without sufficient copies of the book for each student, or for all students when comprehension of critical passages in a book must be reinforced, is partner or team reading. Students reread the selection aloud with a partner, trading the roles of reader and listener page by page. Because the roles are traded so frequently, both stay engaged in the reading. By

pairing more and less proficient readers, partner reading provides the opportunity for peer correction and guidance. The average to "challenged" reader can benefit tremendously when given frequent opportunities to read multiple paragraphs in an uninterrupted manner. This also supports the self-monitoring of comprehension by independent readers.

After completing the reading selection, student teams first discuss answers to questions about the reading. Group discussions are held with pencils down so that students listen and engage in the discussion rather than work on answering the questions and thereby bring answers to a higher level. The partner or team discussion of these questions allows students to cooperatively manage the discussion and practice important social skills while the teacher circulates from team to team to monitor the discussion, guiding groups back on track if they have strayed, and ensuring that the group discussion successfully guides readers to arrive at appropriate responses.

After group discussion, students write out answers individually. The questions provide an opportunity to practice writing skills and refer back to the text, thus illuminating the process of building responses. The open-ended questions ask students to summarize what they have just read, recall details of a passage, connect it to prior reading, predict what might happen subsequently, and synthesize and interpret the reading in their own words. The questions are written in such a way as to unlock the key ideas in the reading selection. Like the pre-reading activities, these questions are designed to develop critical thinkers who are constantly negotiating with and bringing meaning to print.

Prior to a formal assessment, students review the reading by answering questions, making predictions about what lies ahead in the novel or non-fiction selection, and complete a writing activity related to the reading selection. This review and the daily ungraded assessment activities provide ample opportunity for all students to study for and clearly understand what is expected of them on the graded test. The review questions are the same as or similar to questions that will appear on the literature test. These informal assessments provide a low-stakes arena in which the teacher can judge a student's progress and target specific areas for attention. Students gain practice taking tests and have expectations clearly delineated prior to formal assessments. Complementing these review activities is the explicit modeling of study and note-taking skills, such as the construction of graphic organizers to summarize main ideas or compare and contrast ideas.

Effective teacher implementation of Student Team Literature is facilitated by extensive training offered for graduate-level university credit in the Philadelphia pilot sites during both the summer and the school year. These professional development activities consider the teacher to be a learner and employ carefully crafted, interactive lessons which include

modeling and guided practice in the implementation of the recommended curriculum and instructional practices. Implementation is also supported by classroom visits by, and team-teaching opportunities with, instructional facilitators or in-school curriculum coaches with subject-specific expertise. This two-pronged approach to training, with initial training outside the classroom followed by classroom visits, supports the successful implementation of the curriculum and instructional practices in the actual classroom.

For each of over 100 literary works, the available Partner Discussion Guides contain similar types of exercises. For example, each features the writing of meaningful sentences to learn new vocabulary words, so that skills learned during the study of previous novels are practiced and reinforced later in the year. Indeed, the four core subject areas of the Talent Development curriculum (language arts, mathematics, science, and social studies) employ the same learning strategies, bringing another layer of reinforcement and cross-fertilization to the Talent Development model.

Listening Comprehension is a 20-minute period of time during which students are read to by a model reader, their teacher. Students hear a fluent reader read and think aloud about what has just been read, exercise comprehension skills, learn about literary elements and devices that students will encounter in the selections they read in Student Team Literature, and improve listening skills. Listening Comprehension extends and reinforces the reading and comprehension skills used in partner or team reading and individual silent reading.

Talent Development Writing complements Student Team Literature and extends the explicit teaching and practice of writing, critical reading, and editing. It places especially heavy emphasis on several important steps in preparing students for the writing experience (e.g., a teacher thinking aloud and modeling his or her approach to writing, and "springboard activities" intended to spark the creative process). Talent Development Writing also emphasizes important activities situated between the steps of the writing process (e.g., teacher-student pre-writing and editing conferences, and conferences with student partners).

CRESPAR

Johns Hopkins University
Center for Social Organization of Schools
3003 North Charles Street – Suite 200
Baltimore MD 21218
410-516-8800 / 410-516-8890 fax

Howard University
2900 Van Ness Street, NW
Washington DC 20008
202-806-8484 / 202-806-8498 fax